



<https://helda.helsinki.fi>

---

## The location and boundaries of consciousness: a structural realist approach

Loorits, Kristjan

2018-09

---

Loorits , K 2018 , ' The location and boundaries of consciousness: a structural realist approach ' , Review of Philosophy and Psychology , vol. 9 , no. 3 , pp. 523-537 . <https://doi.org/10.1007/s13164-018-0385-4>

---

<http://hdl.handle.net/10138/326185>

<https://doi.org/10.1007/s13164-018-0385-4>

---

unspecified

acceptedVersion

---

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

# The location and boundaries of consciousness: a structural realist approach

Kristjan Loorits<sup>1</sup>

**Abstract:** Despite the remarkable progress made in consciousness research during recent decades, there is still no sign of a general agreement about the location of its object. According to internalists, consciousness resides inside the brain. According to externalists, consciousness is partly constituted by elements or aspects of the environment. Internalism comports better with the existence of dreams, hallucinations and sensory imaging. Externalism seems to provide a more promising basis for understanding how we can experience the world and refer to the content of our consciousness. I argue that the framework of structural realism supports internalism and helps to reveal the reasons behind the apparent explanatory success of the externalist approach. More specifically, structural realism supports the view that the structure of our consciousness is always present in our neural processes and only sometimes (additionally) in an extended system that includes elements of the environment.

## 1. Introduction: The Location Problem

Recent decades have engendered several promising and naturalistic but mutually incompatible explanatory approaches to consciousness. A particularly wide and fundamental disagreement prevails between internalists and externalists about consciousness (not to be confused with related yet distinct internalism-externalism juxtapositions regarding meaning, cognition, representational content etc.). We may call it *the location problem* of consciousness.

According to consciousness internalists, consciousness should be explained in terms of neural processes (e.g. Revonsuo 2006, 2015; Searle 2000; Koch 2004; Clark 2009). Consciousness externalism, on the other hand, comes in many different varieties the most obvious common denominator being the rejection of consciousness internalism.

---

<sup>1</sup> Department of Philosophy, History and Art Studies, University of Helsinki, Helsinki, Finland

The more traditional versions of consciousness externalism have their roots in philosophy of language meaning language). For example, it has been often argued that the content of our consciousness must be public in a sense that neural processes are not. Similarly, it has been sometimes claimed that the phenomenal content of consciousness is *wide* in Putnamian sense and thereby partly constituted by external objects (e.g. Tye 1995; Dretske 1995).

The more empirically driven and increasingly popular externalist strategies stress the importance of an For example, according to the so called sensorimotor theory, consciousness should be explained in terms of sensorimotor interactions with the environment or in some closely related phenomena such as sensorimotor skills or sensorimotor knowledge (e.g. Noë 2004, 2009; Ward 2012; Hurley 2010;

I examine the location problem from the point of view of structural realism (by relying mainly on the version of structural realism defended by Ladyman and Ross 2007). One of the most valuable features of the view (presumably first noticed by Worrall 1989) is that it allows us to evaluate and compare the accuracy of competing theories while ignoring their ontological commitments (except the ontological commitment to the data; see Ladyman and Ross 2013). In other words, structural realism allows us to put ontological and purely conceptual considerations aside and compare rival theories by focusing solely on their mathematical structures the more accurately the structure of a theory matches the structure of the explanandum, the more accurate the theory. Of course, a mere structurally accurate description of an explanandum does not pass as a theory the structure should be also discovered (or discoverable according to the theory) as a relational structure in some empirical data.

In a nutshell, according to structural realism, internalism and externalism about consciousness would be genuinely incompatible only if (and as far as) they fail to provide us with the same empirically testable structural description of consciousness. Therefore, from the point of view of structural realism, the mere fact that different approaches search for consciousness from different locations does not imply that those approaches must be necessarily incompatible. For it could happen that the very same relational structure is found in the data collected from neural processes and from the environment.

For example, if a healthy and awake subject stares at a slowly flashing red light, then the *temporal* structure of her experience of redness matches the temporal structure of the activity of her neural correlate of the experience of redness. At the same time the very same temporal structure matches also the temporal structure of the flashing of the actual red light in her

experience of redness can be found in at least two different locations (in the brain and in the environment).

Of course, discovering one simple structural aspect of consciousness in two different locations says very little about the location of consciousness as a whole. Nevertheless, it says something: discovering some structural aspect of consciousness in one location does not rule out the possibility that the very same structural aspect can also be found in some other location.

Therefore, in light of structural realism, if two theories disagree about the location of the phenomenon (or about the location of the explanatory basis of the phenomenon), then, in order to establish the superiority of one theory over the other, one must demonstrate at least two things: first, that some of the structural aspects found in location L1 could never be found in a competing location L2; and second, there must be no (known) principled reasons why some of the structural aspects of the phenomenon could not be found in L1. If these two conditions are met, then it can be argued that from the two locations under consideration, only in L1 can one hope to find *all* structural features of the phenomenon (i.e. the entire structure of the phenomenon).

As the above example shows, some simple structural aspects of consciousness can be easily found both in the brain and in the environment. However, it can be argued that certain structural interactions with the environment (the most obvious examples being the structures of dream experiences of permanently paralyzed people). On the other hand, despite consciousness externalists' efforts to show otherwise, there seems to be no compelling evidence to support the idea that some structural aspects of consciousness could never be found in neural processes. Therefore, as I argue in section 3, the internalist brain-bound view is a more promising candidate for a general theory of consciousness.

Another contribution of structural realism concerns the externalist ideas that originate from philosophy of language. It has been often argued, in the spirit of Wittgenstein, that phenomenal content of consciousness must be public in a sense that neural processes are not. For the communication about the phenomenal content is carried out entirely in terms of public phenomena and does not se

However, in light of structural realism, internalism about consciousness is fully compatible with the idea that the structures of our *veridical* perceptual experiences can be found in the environment or in our interactions with the environment. If so, then while describing the content of my consciousness, I would describe in most of the cases (i.e. in the cases of veridical

perceptual experiences) some perfectly public relational structures that are accessible to all the members of my linguistic community. But if the brain-bound view is correct, then those very same relational structures are also present in my neural processes, and hence I would describe, in a sense, also those neural processes (by giving an accurate structural description of those). However, when describing my dreams or hallucinations, I would describe certain relational structures that can be found only in my neural processes, although I would describe them by using the vocabulary of public phenomena. Therefore, as I argue in section 4, the framework of structural realism allows us to combine internalism about consciousness with the popular idea according to which the possibility of communication depends always on the publicity of meaning.

Similarly, in light of structural realism, consciousness internalism is not threatened by the idea that the content of our consciousness might be wide in Putnamian sense. For if consciousness internalism is only committed to the view that the *structure* of our consciousness can be found in neural processes, then it does not have to fight the idea that the language we use to describe that structure depends on the wide meanings of its words (see section 4 below).

The approach I propose is novel, for the potential of structural realism as a framework for evaluating and comparing rival theories of consciousness has remained largely unexplored. In general, the research on possible mutual relevance of structural realism and consciousness research is only just beginning. There are some studies focusing on the challenges of reconciling structural realism with the existence of apparently non-structural qualia (e.g. Unger 2001; Lloorits 2014). Relations between contemporary versions of structural realism and Russellian monism in the context of consciousness have also been examined (e.g. Ney 2015).

It should be noticed that my primary goal is not to defend the framework of structural realism, but to examine how (and if) its core ideas could help us unravel the location problem of consciousness. Therefore, I will not provide any detailed arguments for structural realism, but merely introduce some of its central theses that are relevant for the location problem (section 2). In the last section (section 5) I examine some specific challenges that the structuralist approach to consciousness must face.

## **2. Structural Realism**

According to the core idea of structural realism, the only scientifically significant aspect of whichever scientific theory is the mathematical structure of that theory (Worrall 1989). All the

rest (self-subsistent individual particles, objects, fields, forces, etc.) is, according to the view, merely a collection of imaginary items that help scientists to think about their theory.

The main motivation for structural realism has its roots in a debate about scientific realism. A well-known virtue of scientific realism is that it helps us understand why science is successful: if entities postulated by our best scientific theories (particles, fields, forces etc.) are real, then the success of novel predictions based on the existence of these entities is understandable. In contrast, the rejection of scientific realism would make the success of science look miraculous (Putnam 1975: 73). Hence the above argument for scientific realism is often called *the no-miracles argument*.

On the other hand, since most of the entities postulated by our once successful past scientific theories are now considered as nonexistent (ether, phlogiston, gravitational force etc.), we have a reason to believe that the entities postulated by our best current scientific theories will be someday abandoned as well. The above argument is called *the pessimistic meta-induction*.

According to structural realism, the mathematical structures of our best scientific theories describe the real structure of the world – and that suffices to explain why science is successful. And since significant parts of the mathematical structures of our best scientific theories from the past have survived the radical theory changes, the pessimistic meta-induction is not a threat to structural realism. For example, there is a clear structural continuity between the theory of Newton and the one of Einstein (e.g. Worral 1989).

Besides, it is widely accepted since Russell (1927) that all our empirical knowledge about scientific objects is limited to their structural-relational properties. For example, all we can ever know about electrons concerns their relations with other particles and objects (including some measuring apparatus). In other words, every single property of an individual electron (its mass, charge, spin, momentum etc.) can be ultimately fully analyzed in terms of such relations. Therefore, our theories about electrons describe certain relational structures which are the only aspects of electrons that can be empirically accessed and experimentally tested.

Although most structural realists focus on the theories of fundamental physics, the framework has been shown to be applicable also to the theories of special sciences. For example, according to Ladyman and Ross (2007), the genuine objects of *all* sciences (i.e. fundamental physics and special sciences alike) are the so called *real patterns* (the notion is borrowed from Dennett 1991, but developed significantly further), which are basically patterns in a data that must satisfy certain criteria specified in information-theoretic terms (see section 5 below). And although each individual theory would have to treat some of its most basic data

elements as a sort of non-structural given, such data elements would be, according to Ladyman and Ross (2013), analysable in fully structural terms in the context of some other theory.

According to Ladyman and Ross (2007), there are two main reasons why real patterns should be considered objectively real and not merely useful explanatory tools in an instrumentalist sense: first, all real patterns must lead to novel predictions (just like real unobservable entities in the framework of scientific realism), and second, every real pattern must be maximally efficient in a sense that there must be no other physically possible patterns (i.e. patterns that are physically possible to discover) that would perform its explanatory and predictive functions more efficiently (in information-theoretic sense). Thereby the reality of a pattern becomes an entirely objective matter that is empirically testable and falsifiable. It also follows that structural realism makes a clear distinction between merely explanatorily useful patterns (which are accepted in an instrumentalist sense, and which may often approximate real patterns) and real patterns (which are difficult to find and whose existence is independent of our knowledge of them).

It should be also acknowledged that Ladyman and Ross have no problem with the causal (or mechanistic) explanations that dominate in special sciences. Although they deny the fundamental metaphysical status of causality, they admit that causal explanations are *indispensable* in special sciences. According to their view, if things go well, then the theories of special sciences by book-keeping them as individuals view allows multi-level mechanistic explanations that are currently influential in cognitive neuroscience (see Craver 2007).

At the most general level, structural realism can be divided into epistemic and ontic versions (see Ladyman 2014), but for the present purposes it is not necessary to pick a favorite. It is enough to follow the most general idea shared by both factions: the only empirically accessible elements in our scientific theories are certain relational structures. While ontic structural realists deny explicitly the existence of anything but structures, epistemic structural realists postulate *unknowable* non-structural relata that realize the knowable relational structures. In any case, both parties agree that specific ontological content of scientific theories (e.g. waves, particles, forces, fields, matter) should not be considered real in any metaphysical sense.

Of course, scientists are free to posit all kinds of non-structural entities as heuristic devices or intellectual tools as long as they do not think of them as metaphysically real. Simply put, according to structural realism, the curvature of spacetime is no more real than the gravitational

force, even though the former is proven to be a more sophisticated tool for thinking about gravity, which is (according to structural realism) ultimately nothing but a certain relational structure. Also, the superiority of one intellectual tool over the other is not a trivial matter. For example, it could be argued that the more general and coarse aspects of gravity are easier to understand in terms of forces, but that the fine details become comprehensible only if approached as aspects of curvature of spacetime. Similarly, it is possible that different structural aspects of consciousness are easier to understand in terms of different kinds of phenomena.

The reason why we are, nonetheless, entitled to say that gravity is rather the curvature of spacetime than a force is that the idea of the curvature of spacetime allows us to think about gravity more accurately. In other words, it allows us to formulate a more accurate *informal description* of gravity. Similarly, even though the externalist approach might help us grasp many general aspects of consciousness more easily, the internalist approach has a potential for providing more accurate account of the phenomenon, or so I argue. In addition, unlike in case of gravity, the rival views of consciousness locate their object (or its explanatory basis) in different places. Hence, if we are trying to solve the location problem, we should obviously prefer the view that gets the location right. And in light of structural realism, the location of a phenomenon should be understood as a location where one can gather data in which the structure of the phenomenon can be found.

### **3. Internalism vs. Externalism in Light of Structural Realism**

As already argued in the introduction, in light of structural realism, the mere positive success of finding certain structural aspects of consciousness in some natural phenomena (e.g. in neural processes or in the environment) is not enough for establishing the superiority of the corresponding view. For it is possible that the structure found in one location can also be found in some other location. Therefore, in order to establish the superiority of one view over the other, it should be demonstrated that some structural aspects found in one location cannot be found in the competing location for some principled reasons. And indeed, the efforts to demonstrate just that are made by both externalists and internalists about consciousness.

In a nutshell, defenders of consciousness internalism like to point out that there are plenty of conscious experiences that do not correspond to anything in the environment or in our interactions with the environment (dreams, hallucinations, unexpressed thoughts, feelings, sensory imaginings, etc.). Consciousness externalists, on the other hand, tend to argue that



certain complex phenomenal features of consciousness cannot be explained in neural terms (e.g. qualia or raw feels, shape of an egocentric space, active and attentional character). However, in light of structural realism, the accusations made by internalists seem to stand on a much firmer ground than the ones made by externalists. In the words of Tononi and Koch (2008: 240):

We are conscious of our thoughts, which do not seem to correspond to anything out there; we can also imagine things that are not out there. When we do so, sensory areas can be activated from the inside (Kosslyn et al. 2001), though there are some differences (Amedi et al. 2005). Also, stimulus-independent consciousness is associated with its own patterns of activation within cortex and thalamus (Mason et al. 2007). During dreams, we are virtually disconnected from the environment (Hobson et al. 2000) – hardly anything of what happens around us enters consciousness, and our muscles are paralyzed (except for eye muscles and diaphragm). Nevertheless, we are vividly conscious: all that seems to matter is that the thalamocortical system continues to function more or less as in wakefulness, as shown by unit recording, EEG, and neuroimaging studies performed during rapid eye movement

evidence also indicates that neither sensory inputs nor motor outputs are needed to generate consciousness. For instance, retinally blind people can both imagine and dream visually if they become blind after 6–7 years of age or so (Hollins 1985; Buchel et al. 1998). Patients with the locked-in syndrome can be almost completely paralyzed, and yet they are just as conscious as healthy subjects (Laureys et al. 2005) and can compose eloquent accounts of their condition (Bauby 1997).

The above kind of evidence is emphasized by many critics of consciousness externalism, and has been used explicitly against the currently popular sensorimotor approach by Revonsuo

correlate with certain neural processes in the absence of any corresponding external factors are hallucinations and dreams of permanently paralyzed people. Simply put, if a permanently paralyzed person has a vivid and detailed dream about a green goblin, there is nothing in her environment or in her behavior where one could find the structure of her dream. However, there is evidence that certain neural processes correlate with the content of our dreams (see Maquet et al. 1996; LaBerge 2000; Horikawa et al. 2013; Siclari et al. 2017). And following the example from the introduction, those neural correlates can be interpreted as certain structural aspects of the corresponding consciousness. For the very least, most of the neural correlates share with

the corresponding aspect of consciousness their temporal structure (assuming the correlations are strong enough). For instance, the neural processes that correlate (strongly) with my experience of redness would occur *every time* I experience red. Therefore, the internalist approach allows us to detect and discover temporal structures of different aspects of perceptual experiences and dreams alike. On the other hand, it would be impossible to detect such specific environment or in her sensorimotor interactions with the environment.

Some proponents of the sensorimotor theory try to avoid the problem by shifting the explanatory burden from the actual interactions to sensorimotor skills or sensorimotor dispositions (2004; Myin 2016; Beaton 2013). According to the general idea, the phenomenal content of consciousness is determined by the specific ways the conscious being is *disposed* to interact with the environment. For example, when seeing a tomato, a subject might be disposed to reach for it, to report seeing it, to try to bite it, etc. But more importantly, when merely hallucinating or dreaming about a tomato, her dispositions would (arguably) be the same. Of course, while dreaming about a tomato, the subject cannot actually move because of so-called *REM atonia* (a state of paralysis of most voluntary muscles), but as Beaton (2013) argues, the subject *could* move if she were not paralyzed, and thereby the relevant dispositions remain intact.

However, in light of structural realism, the above strategy has the following problem: although it might offer a method to *describe* and *analyze* all experiences in non-internalist (non-neural) terms, it does not specify where could one find the specific structures of individual dreams if not in neural processes. And if those structures are to be found only in neural processes (as some sensorimotor theorist are willing to do), then we are back in the internalist framework regardless of the fact that the explanatorily relevant neural structures can be *described* by using the vocabulary of the sensorimotor theory (see section 4 below).

Summing up, while neither the externalist nor the internalist theories can currently explain dreams, internalism offers at least an empirical prediction of where would the structures of individual dreams be eventually found (i.e. where can one gather data in which the structures of individual dreams can be found). And the problem with the externalist approach is precisely that it fails to offer any such prediction.

However, as mentioned earlier, externalists are prone to argue that certain complex phenomenal features of consciousness can be never explained in terms of neural processes. According to their accusations, internalism is committed to the wrong kinds of concepts. For example,

*feels* of experiences (the so-called *qualia*)

concepts and

language we can use to describe the modes of interaction [are] the same concepts and language

Thompson (2004) argue that the internalist project of tracking the neural correlates of consciousness is inescapably committed to the concept of *receptive field* which is, according to them, inappropriate for understanding and explaining certain complex phenomenal features of consciousness (such as the shape of an egocentric space, the active and attentional character, the structural coherence of phenomenal content, etc.).

However, in light of structural realism, the only thing the brain-bound view is truly committed to, is the idea that the *structure* of consciousness can be found in neural processes. The question of which vocabulary, language or concepts should or could be used to describe that structure is a matter of practical convenience.

In its standard use, the concept of receptive field refers to a property of a single neuron: its capacity to respond to a particular kind of external stimuli that might *or might not* be consciously experienced (I thank an anonymous referee for pointing that out). On the other hand, the concept of *neural correlate of consciousness* refers to a neural activity (that might be complex and involve large neuronal assemblies) that systematically co-occurs with consciousness or with some specific content of consciousness. Admittedly, the two notions are related. For example, the neural correlate of experiencing movement in a particular area of the visual field would involve neurons in higher visual cortex that respond to moving stimuli in that particular area of the visual field.

So, the concept of receptive field might be sometimes useful for specifying the nature of some neural correlates of consciousness. And keeping in mind that the (strong) neural correlates of consciousness share with the corresponding aspect of consciousness (at least) their temporal structure, it can be also said that the concept of receptive field might be sometimes useful for describing some simple structural aspects of consciousness in terms of neural phenomena. But it does not follow that the concept should (or could) be used to describe more complex structural aspects of consciousness and more importantly, it does not follow that those complex structural aspects cannot be found in neural processes.

In light of structural realism, in order to demonstrate that the complex phenomenal features of consciousness are out of the explanatory reach of the internalist approach, externalists ought to demonstrate that the brain lacks sufficient complexity to host the structures of those features. And none of the externalist arguments is designed to do that. The fact that the structures of

complex phenomenal features have not yet been discovered in neural processes does not imply that they are not there to be found. Brains are notoriously complex, and most of the higher order neural structures are yet to be discovered – and that suffices to explain why we cannot currently explain the complex phenomenal features in terms of neural processes. In contrast, sleeping sensorimotor interactions with the environment are extremely simple (by being virtually nonexistent), and thus there is no hope that we could one day explain the rich and he environment. And as seen above, appealing to sensorimotor skills or sensorimotor dispositions does not help, for those notions fail to lead to any predictions of where could one find the specific structures of individual dreams.

#### **4. The Legacy of Philosophy of Language**

One enduring argument against the internalist approach is based on the idea that, in a sense, the nature of our neural processes is irrelevant for the identification of consciousness. It seems to be a fact that people can talk about consciousness – both about their own and the ones of the others. However, while I can meaningfully claim that my friend is conscious, I have (usually) no epistemic access to her neural processes. In the words of G. H. Von Wright (1998: 134):

And as the above is true about mental phenomena in general, it is also true about consciousness in particular.

So, the core of the argument is the following: the environment and our interactions with the environment are perfectly public phenomena while the processes inside our brains are (usually) hidden from us. Therefore, while we talk about consciousness, we must be talking about those public phenomena, for we simply cannot be talking about neural processes to which we have neither perceptual nor epistemic access (except in some rare and artificial laboratory conditions, which are not the paradigmatic context of our daily references to consciousness).

Relatedly, it has been sometimes argued that the phenomenal content of our consciousness is *wide* in Putnamian sense, being partly constituted by the facts about the external world (e.g. Tye 1995; Dretske 1995). In principle, the above claim can be made about veridical experiences and dreams alike. Let us say I dream about a blue cow. Then it can be argued that although my experience of being near a blue cow is not veridical, it has, nevertheless, objective and external

truth-conditions (i.e. me being near a blue cow) that constitute partly the *meaning* and thereby also the content of my dream.

However, from the perspective of structural realism, the above language- and meaning-based arguments do not undermine consciousness internalism. For according to structural realism, when we talk about the location of consciousness, we talk about location(s) where one can gather data in which the structures of individual experiences can be found. And in the case of veridical perceptual experiences, we talk indeed about some perfectly public and easily accessible structures. However, if the brain-bound view is correct, then the structures of our veridical experiences are *also* present in our neural processes. Consequently, when we describe our veridical experiences to the others, we describe relational structures that are simultaneously present in public external phenomena and in our neural processes. And if it is true that the possibility of communication depends on shared and public meaning, then it is also understandable why we describe and conceive *all* our conscious experiences in terms that have shared and public meaning – it seems to us that we dream about green goblins and blue cows and not about neural activity patterns. Yet the structure of our dream experience exists only in our brain, although we describe that structure (both to the others and to ourselves) by using the vocabulary of public and shared phenomena.

Similarly, even if the truth-conditions that specify the meaning and the content of my dreams are public and external, what matters in the light of structural realism, is that the detailed structure that individuates some particular dream and distinguishes it from all the other phenomena can be found only in my neural processes. Therefore, although we rejected the externalist idea that dreams can be *explained* in terms of sensorimotor skills or dispositions (or any other external phenomena), we may accept the idea that dreams are complex neural processes whose structures can be *described* by using the vocabulary of sensorimotor skills or dispositions (or some other external phenomena).

It is also significant that in some situations neural findings are relevant for identifying consciousness while behavior is not. There is a famous case of a patient, who was diagnosed with vegetative state but was later discovered to be conscious because she was capable of following the verbal instructions of the researchers, who asked her to imagine either playing tennis or walking through her room (Owen et al. 2006). Her success at those imaginary tasks was verified by fMRI. Later experiments have produced similar results (e.g. Monti et al. 2010; Naci and Owen 2013; Cruse et al. 2011), and according to current estimation, up to 20 per cent of vegetative patients are actually conscious.

## 5. Some Challenges for the Structuralist Approach

One of the background assumptions of the structuralist approach outlined in this paper has been that it is meaningful to claim and reasonable to hope that the structure (the entire structure!) of consciousness will be someday found in neural processes. The idea in itself is not new, and it seems to be endorsed (either tacitly or explicitly) by many proponents of the internalist approach. For example, Revonsuo (2006) believes that the structure of consciousness will be discovered in neural processes once we learn to monitor the *proper level of organization* of neural activity – in any other level we would find only neural correlates of consciousness. In other words, in those lower levels we would find patterns of neural activity that systematically co-occur with some phenomenal aspects of consciousness. But only in the proper level of organization would we find a complex pattern that simply *has* the structure of consciousness. Obviously, the above way of thinking presupposes that there is some objective and natural way for establishing structural similarity and structural identity (or structural isomorphism) between different kinds of phenomena. And some externalists deny such possibility. For example, according to which the similarity and difference relations between sensory qualia correspond to the similarity and difference relations between corresponding neural processes. According to it is not at all natural to say that an experience of redness is more

there

There is evidence that color hues are encoded in visual area V4 by adjoining patches of neurons, so that neighboring patches that correspond to similar colors overlap considerably (e.g. Conway and Tsao 2009; and especially Li et al. 2014). Put simply, the bigger the phenomenal similarity between any two colors, the larger the overlap of the patches that encode them. Thus, neuronal assemblies that encode red and orange share more neurons than the assemblies that encode red and yellow, whereas the assemblies that encode red and blue share none. The above findings suggest that the brain state corresponding to a certain color is not just a simple state, but a complex state that is structured in a way that corresponds to the structure of the color it represents. This way of ordering the brain states so that their similarities correspond to perceptual judgments about similarities between colors [...], why is it this way of ordering the brain states, rather than

similarity presupposes some method of projection, and two structures that are very similar under one method of projection might be very dissimilar under another. Thence the worry: if there is then the whole notion of structural similarity (or structural isomorphism) threatens to become hopelessly relativistic.

But one may now ask: if the structural similarity is problematic in such a fundamental and general sense, is there any other kind of similarity that is not? For example, if we agree that a lion is more similar to a tiger than to a rock, what kind of similarity we have in mind? Or what do we mean by saying that the theory of gravity of Einstein is more similar to the theory of gravity of Ne

bodies move toward their natural place)? Majority of philosophers would agree that all the properties that determine the similarity between two objects or between two theories can be analyzed in fully structural terms. The philosophical questions about the similarity relation and its dependency on methods of projection are interesting in their own right, but since scientists keep discovering and acknowledging all kinds of similarities that can be analyzed in fully structural terms, we should not be too worried about the legitimacy of the idea of structural similarity between different neural states.

Moreover, the framework of structural realism might actually provide us with the necessary tools for evaluating and comparing different methods of projection. As already noted, according to structural realism, science is in the business of gathering data and finding structural patterns in that data. According to the view, the structural patterns thus found are the very structures we ought to be realists about. But obviously, not every relational structure discovered (or discoverable) should be considered as a genuine scientific object. Consider, for example, the

criteria the relational structure must satisfy in order to pass as a genuine scientific object.

The most important of such criteria is, according to Ladyman and Ross (2007), the requirement that the structural pattern must allow us to *encode information* in the data more efficiently, in information-

words, the structure found in the data must *compress* the information in the data.

Now, following the lines of the above analysis, the structural similarity between two phenomena can be seen as a structural pattern itself. In a nutshell, two phenomena are

structurally similar if (and only if) there is a way to describe their structures together more cost-efficiently (in information-theoretic sense) than describing both structures separately. For example, if there is a method of projection that allows us to describe two brain states together more cost-efficiently than describing both of them separately, then the two brain states are structurally similar in an objective sense. Also, if there is a method of projection that allows us to describe phenomenal consciousness and certain neural processes together more cost-efficiently than describing both of those phenomena separately, then we are dealing with a genuine structural similarity. Correspondingly, the method of projection that allows us to describe the two phenomena together as cost-efficiently as physically possible should be considered as the best method of projection in an objective sense.

Structural identity, on the other hand, would require the existence of a structure that compresses the information about two (or more) phenomena with *maximal* efficiency, so that the structural description of one of the phenomena could be used to describe also the other(s). In any case, what is important for the present purposes is that there are some perfectly objective criteria for comparing and evaluating different methods of projection and those criteria might allow us to single out

Last but not least, some philosophers argue that finding the structure of consciousness in neural processes is not enough because consciousness is something over and above its structure. For example, David Chalmers, the author of the so called *hard problem of consciousness*, agrees that it is probable that the structure of consciousness is identical with some informational structure in our brains, but that consciousness has besides its structural properties also qualitative properties known as *qualia* (Chalmers 1995, 2003).

Indeed, the problem of qualia may turn out to be the single biggest challenge for structural realism as a general metaphysical framework. However, it is worth acknowledging that recent years have introduced some hypotheses according to which qualia might be fully structural. For example, Pereboom (2011) argues that qualia might be compositional, but appear in introspection as primitive and monadic. A more empirically based neurobiological account has been put forward by Crick and Koch (1998), who suggest that qualia are complex networks of unconscious associations (see also Koch 2004; Loorits 2014). For the record, structural accounts of qualia have been also put forward by externalists. For example, according to

analyze or break down into the experience of atomic elements, or simple features. Experience



The above approaches resonate well with structural realism. But considering their tentative and speculative nature, it is premature to hope that a fully structural and explanatorily satisfactory account of consciousness is around the corner (or ever developed). More generally, it is difficult to understand how consciousness could be, by its most fundamental metaphysical nature, nothing but a complex pattern in some data. Therefore, let us take seriously the possibility that a complete theory of consciousness must exceed the boundaries of structural realism (thus posing a serious threat to the latter as a general metaphysical view of reality).

However, even if structural realism failed to provide us with a metaphysical account of consciousness, it remains still a valuable analytical framework for approaching the location problem. For *none* of the defenders of the irreducibly non-structural view of qualia denies that some properties of consciousness are structural (see Chalmers 1995, 2003; Unger 2001). Consequently, acknowledging the possibility that consciousness is something over and above its structure, the structural features of consciousness are still relevant with respect to the location problem. Simply put, consciousness cannot be explained by a theory that fails to explain some of its structural properties.

Summing up, the main contribution of structural realism with respect to the location problem is that it provides us with a single framework that allows us to compare and evaluate theories with very different background assumptions and ontological commitments. More specifically, structural realism allows us to organize and evaluate different types of arguments for and against those theories: what matter most are neither the positive results of finding some structural aspects of consciousness in some definite location (for the same structure might be present in another location as well), nor the particular failures to find certain structural aspects in some location (for those might be found there in the future), but the arguments that demonstrate why certain structural aspects of consciousness could never be found in certain locations. And regardless of the possible developments in the research on qualia, it seems that consciousness (or its explanatory basis) cannot reside in a phenomenon that lacks some of its structural properties. In other words, whatever consciousness turns out to be, it must be something that has all its structural properties.

**Acknowledgements:** This research has been supported by a scholarship from the Finnish Cultural Foundation.

## References

- Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. *Neuron*, 48, 859–872.
- Bauby, J. D. (1997). *The diving-bell and the butterfly: A memoir of life in death*. New York: Alfred A. Knopf.
- Beaton, M. (2013). Phenomenology and Embodied Action. *Constructivist Foundations*, 8(3), 298-313.
- Buchel, C., Price, C., Frackowiak, R. S., & Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain*, 121(3), 409–419.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- Chalmers, D. (2003). Consciousness and its place in nature. In Stich, S. & Warfield, F. (Eds.), *The Blackwell Guide to Philosophy of Mind* (pp. 247–272). Oxford: Blackwell.
- Clark, A. (2009). Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head. *Mind*, 118, 963–993.
- Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Sciences of USA*, 106, 18034–18039.
- Craver, C. F. (2007). *Explaining the Brain*. New York: Oxford.
- Crick, F., & Koch, C. (1998). Consciousness and Neuroscience. *Cerebral Cortex*, 8, 97–107.
- Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T. A., Fernández-Espejo, D., et al. (2011). Bedside detection of awareness in the vegetative state: a cohort study. *Lancet*, 378, 2088–2094.
- Dennett, D. (1991). Real patterns. *Journal of Philosophy*, 88, 27–51.
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Hobson, J. A., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences*, 23, 793–842; discussion pp. 904–1121.
- Hollins, M. (1985). Styles of mental imagery in blind adults. *Neuropsychologia*, 23, 561–566.
- Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, 340, 639–642.
- Hurley, S. (2010). Varieties of Externalism. In Menary, R. (Ed.), *The Extended Mind* (pp. 101–154). Aldershot: MIT.

- Koch, C. (2004). *The quest for consciousness: a neurobiological approach*. Englewood: Roberts and Co.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2, 635–642.
- Ladyman, J. (2014). Structural Realism. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Available via DIALOG. <http://plato.stanford.edu/entries/structural-realism/> Retrieved 20 Jan 2017
- Ladyman, J., & Ross, D. (with Spurrett, D., & Collier, J.) (2007). *Every Thing Must Go: Metaphysics Naturalised*. Oxford: Oxford University Press.
- Ladyman, J., & Ross, D. (2013). The World in the Data. In Ross, D., Ladyman, J., & Kincaid, H. (Eds.), *Scientific Metaphysics* (pp.108-150). Oxford: Oxford University Press.
- LaBerge, S. (2000). Lucid dreaming: Evidence and methodology. *Behavioral and Brain Sciences*, 23, 962–964.
- Laureys, S., Pellas, F., Van Eeckhout, P., Ghorbel, S., Schnakers, C., et al. (2005). The locked-in syndrome: What is it like to be conscious but paralyzed and voiceless? *Progress in Brain Research*, 150, 495–511.
- Li, M., Liu, F., Juusola, M., & Tang, S. (2014). Perceptual Color Map in Macaque Visual Area V4. *The Journal of Neuroscience*, 34(1), 202–217
- Loorits, K. (2014). Structural qualia: a solution to the hard problem of consciousness. *Frontiers in Psychology*, 5(237).
- Maquet, P., Péters, J., Aerts, J., Delfiore, G., Degueldre, C., et al. (1996). Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature*, 383, 163–166.
- Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., et al. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, 315, 393–395.
- Monti, M. M., Vanhaudenhuyse A., Coleman M. R., Boly M., Pickard J. D., Tshibanda L., et al. (2010). Willful modulation of brain activity in disorders of consciousness. *The New England Journal of Medicine*, 362, 579–589.
- Myin, E. (2016). Perception as Something We Do. *Journal of Consciousness Studies*, 23(5-6), 80-104.
- Naci, L. & Owen, A. M. (2013). Making Every Word Count for Nonresponsive Patients. *JAMA Neurology*, 70(10),1235-1241.

- Ney, A. (2015). A Physicalist Critique of Russellian Monism. In Alter, T., & Nagasawa, Y. (Eds.), *Consciousness in the Physical World: Essays on Russellian Monism* (pp. 346-369). New York: Oxford University Press.
- Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.
- Noë, A. (2009). *Out of Our Heads*. New York: Hill & Wang.
- Noë, A., & Thompson, E. (2004). Are There Neural Correlates of Consciousness? *Journal of Consciousness Studies*, 11(1), 3–28.
- O'Regan, K. (2011). *Why Red Doesn't Sound Like a Bell: Explaining the Feel of Consciousness*. New York: Oxford University Press.
- Review of Philosophy and Psychology*, 3, 89-108.
- Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., et al. (2006). Detecting awareness in the vegetative state. *Science*, 313, 1402.
- Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.
- Putnam, H. (1975). *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.
- Revonsuo, A. (2006). *Inner Presence: Consciousness as a Biological Phenomenon*. Cambridge: The MIT Press.
- Revonsuo, A. (2015). Hard to See the Problem? *Journal of Consciousness Studies*, 22(3–4), 52–67.
- Ross, D. (2000). Rainforest realism: A Dennettian theory of existence. In D. Ross, A. Brook, and D. Thompson (Eds.), (pp. 147–68). Cambridge, MA: MIT Press.
- Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul. Reprinted 1954, London: Allen and Unwin.
- Searle, J. (2000). Consciousness. *Annual Review of Neuroscience*, 23, 557–578.
- Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., et al. (2017). The neural correlates of dreaming. *Nature Neuroscience*, 20(6), 872–878.
- Tononi, G., & Koch, C. (2008). The Neural Correlates of Consciousness - An Update. *Annals of the New York Academy of Sciences*, 1124, 239–261.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Unger, P. (2001). Why there are no people. *Midwest Studies in Philosophy*, 4, 177–222.

- Von Wright, G. H. (1998). *In the shadow of Descartes*. Dordrecht: Kluwer Academic.
- Ward, D. (2012). Enjoying the Spread: Conscious Externalism Reconsidered. *Mind*, 121, 731–751.
- Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica*, 43, 99–124.